

How much spatial information is lost in the sensory substitution process? Comparing visual, tactile, and auditory approaches

Article (Accepted Version)

Richardson, Michael, Thar, Jan, Alvarez, James, Borchers, Jan, Ward, Jamie and Hamilton-Fletcher, Giles (2019) How much spatial information is lost in the sensory substitution process? Comparing visual, tactile, and auditory approaches. *Perception*, 48. pp. 1079-1103. ISSN 0301-0066

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/85557/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

How much spatial information is lost in the sensory substitution process? Comparing visual, tactile, and auditory approaches.

Richardson, M.¹, Thar, J.², Alvarez, J.¹, Borchers, J.², Ward, J.^{1,3}, & Hamilton-Fletcher, G.^{1,4,*}

¹Department of Psychology, University of Sussex, Brighton, UK

²Media Computing Group, RWTH Aachen University, Aachen, Germany

³Sackler Centre for Consciousness Science, University of Sussex, Brighton, UK

⁴Visual Science and Neuroimaging Lab, New York University Langone Health, New York City, USA

*e-mail: giles.hamilton-fletcher@nyumc.org

Abstract

Sensory substitution devices (SSDs) can convey visuospatial information through spatialised auditory or tactile stimulation using wearable technology. However, the level of information loss associated with this transformation is unknown. In this study novice users discriminated the location of two objects at 1.2m using devices that transformed a 16×8 depth map into spatially distributed patterns of light, sound, or touch on the abdomen. Results showed that through active sensing, participants could discriminate the vertical position of objects to a visual angle of 1° , 14° , and 21° , and their distance to 2cm, 8cm, and 29cm using these visual, auditory, and haptic SSDs respectively. Visual SSDs significantly outperformed auditory and tactile SSDs on vertical localisation, whereas for depth perception, all devices significantly differed from one another (visual > auditory > haptic). Our findings highlight the high level of acuity possible for SSDs even with low spatial resolutions (e.g. 16×8) and quantify the level of information loss attributable to this transformation for the SSD user. Finally, we discuss ways of closing this ‘modality gap’ found in SSDs and conclude that this process is

best benchmarked against performance with SSDs that return to their primary modality (e.g. visuospatial into visual).

Keywords: sensory substitution; spatial perception; hearing; touch; vision

Introduction

Vision provides an abundance of spatial information, helping to localise individual objects, facilitate orientation, and assist navigation within the environment (Urbanski et al., 2008; Michel & Henaff, 2004; Kelly et al., 2008; Epstein, 2008). Relative to those with visual experience (the sighted, late-blind), congenitally-blind individuals who have never experienced rich visual information showcase a range of subtle shifts in spatial processing, including: A lack of automatic integration between tactile and auditory spatial reference frames (Hötting & Röder, 2004; Hötting, Rösler, & Röder, 2004); distorted representations of space, compromising the ability to make judgments based on the Euclidean distance between auditory targets (Gori et al., 2013); contraction of space resulting in an over-estimation of distance (Cappagli, Cocchi, & Gori, 2017; Kolarik et al., 2017); increased difficulty in replicating movement in space (Gori et al., 2017); and finally, show a preference for, and effective use of, egocentric (self-oriented) representations of external space, but a slower processing of allocentric (environment-oriented) representations, which for some but not all situations can result in increased errors (Heller & Kennedy, 1990; Pasqualotto & Proulx, 2012; Iachini, Ruggiero, & Ruotolo, 2014; Corazzini et al., 2010; Ruggiero, Ruotolo, & Iachini, 2012; Ruggiero, Ruotolo, & Iachini, 2018; Vercillo, Tonelli, & Gori, 2018), which in turn, can result in an impaired spatial memory for the congenitally-blind (Pasqualotto et al., 2013). Overall, the experience of vision continues to influence our representation of space even if no visual input is currently available. As such, providing structured spatial information to the fully-blind not only serves as an intervention to assist in daily tasks (Hamilton-Fletcher et al., 2016; Elli, Benetti, & Collignon, 2014; Maidenbaum, Abboud, & Amedi, 2014; Kristjánsson et al., 2016), but it allows users to interpret their remaining senses more accurately (Cappagli et al., 2017), and explore how our metamodal representation of space can develop through additional sensory stimulation (Pasqualotto & Esenkaya, 2016).


One modern approach to extending the sensory world of fully-blind individuals is to use ‘sensory substitution devices’ that translate the sensory properties of one sense (e.g. vision) into another (e.g. hearing or touch). The substituted modality is defined by the specificity of the sensory signal (e.g. only vision experiences colour), and the sensorimotor contingencies involved (e.g. only vision experiences ‘visual occlusion’ – O’Regan & Noë, 2001). If only a subset of these visual signals are experienced, such as only conveying spatial information, and that the sensorimotor contingencies involved are closest to vision rather than veridical hearing or touch, then the information would be described as ‘visuospatial.’ SSDs use consistent mappings between the senses in order to allow the user to decode the pattern of received signals back into a coherent mental image of the world upon which to act. For SSDs converting vision into sound or touch, these received signals can indicate the position, size, shape, colour, and motion of visual objects (Gori et al., 2016; Hamilton-Fletcher & Ward, 2013). Over time, these can even evoke a qualitative shift in user perception, so that sound or touch can allow the user to perceive spatialized patterns of light (Ward & Meijer, 2010; Ortiz et al., 2011), a form of synthetically acquired synaesthesia comparable to vision (Ward & Wright, 2014). SSDs not only have the potential to provide access to new spatial information and shift metamodal representations of space, but interviews with users suggest a range of practical benefits outside of the lab including enhanced safety, independence, and facilitating curiosity of the environment, further encouraging spatial exploration (Hamilton-Fletcher et al., 2016). Visuospatial SSDs tend to fall under two main types of feedback, haptic (the active use of touch), or auditory (the sense of listening), and while we explore specific examples of these SSDs below, wider overviews of SSDs are available (Gori et al., 2016; Hamilton-Fletcher & Ward, 2013).

Haptic (also known as tactile) devices were first popularised in 1969 with the Tactile Vision Sensory Substitution (TVSS) device (Bach-y-Rita et al., 1969). This device used a

television studio camera (which users could manipulate) connected to vibrating solenoids, positioned in the back of a dentist's chair, arranged in a matrix that resembled the camera's pixels. Bach-y-Rita and his research group continued to improve their design throughout the following decades, experimenting with electro-tactile stimulation, and, most notably, leading to the creation of the Tongue Display Unit (Sampaio, Maris, & Bach-y-Rita, 2001); an SSD that converts 2D greyscale images into a spatialized pattern of electro-tactile stimulation on the tongue in real-time. The output is constrained to a 12×12 matrix with users able to reach a visual acuity of up to 20/430 for single objects (for comparison 20/20 is optimal, and 20/200 is the threshold for blindness). More recently the BrainPort has utilised a 20×20 matrix; however, there is no recorded increase in visual acuity from this (Nau, Bach, & Fisher, 2013; Stronks et al., 2016).

Since Bach-y-Rita's work, vibrotactile technology has progressed with the use of three-dimensional (3D) cameras and mobile computer processing. What once required a film crew, many hands and a somewhat unwieldy dentistry chair can now fit into a single-unit wearable vest (Ertan et al., 1998; Wacker et al., 2016). Many different haptic feedback vests have been developed over the past few years (Ertan et al., 1998; Rochlis, 1998; Jones, Nakamura, & Lockyer, 2004; Spanlang et al., 2010), with the VibroVision among the most recent (Wacker et al., 2016). Like most haptic vests, the VibroVision contains an input source, in this case, a three-dimensional sensor to record the necessary visuospatial information; an image processing unit, that converts the visuospatial information into values that drive, finally, a tactile output, through a series of vibrating motors arranged in a 16×8 matrix on the users' abdomen (Wacker et al., 2016).

While visuospatial-into-audio devices were created much earlier (e.g. Noiszewski's Elektroftalm in 1897 – Starkiewicz & Kuliszewski, 1963) they only gained in popularity for scientific research after the introduction of 'the vOICe' (Meijer, 1992). This software

converts 2D greyscale images into sounds played through headphones, utilising stereo playback and time for the horizontal axis (scanning from left to right over time), pitch changes to represent verticality (higher pitches = higher locations), and volume indicating brightness (louder = brighter). In combination this image  would produce a low-pitched tone in the left ear that rises to a high-pitched tone in the right ear. The vOICe is still considered one of the higher resolution auditory devices, with the device typically outputting at a resolution of 176×64 and allowing a perceptual resolution bordering on 20/200 on single object discrimination (Striem-Amit, Guendelman, & Amedi, 2012; Haigh et al., 2013).

More recent SSDs have tended to prioritise 3D information owing to its immediate practicality for locating individual objects, segmenting their shape, and utility for navigation (Fristot et al., 2012; Stoll et al., 2015; Spagnol, Baldan, & Unnthorsson, 2017; Caraiman et al., 2017; Dunai et al., 2013). When objects within the field-of-view of a 3D sensor are sonified in a manner that replicates the perception of sounds emanating from the location of the object, this process is also referred to as a ‘virtual acoustic space’ (González-Mora et al., 1999, 2006; Rodríguez-Hernández et al., 2010; Eckert, Blex, & Friedrich, 2018). One modern incarnation of this is the ‘Synaestheatre’ which converts a depth image from a 3D sensor into realistically spatialized sounds (Hamilton-Fletcher et al., 2016). The sounds are spatialized using head-related transfer function (HRTF); which describes how the spatial positions of the listener and sound source alter the received sounds in terms of inter-aural timing, intensity, as well as distortions created by the head and pinnae (Algazi, Avendano, & Duda, 2001; Potisk, 2015; Kulkarni, Isabelle, & Colburn, 1999; Kistler & Wightman, 1992). The Synaestheatre produces audio that updates in real-time with sounds that are fully 3D spatialized. As such, the Synaestheatre is similar to the VibroVision in that azimuth and elevation are conveyed spatially (rather than through abstract codes), albeit through hearing and touch respectively.

To date, studies examining how effectively visuospatial information can be converted into hearing or touch have been difficult to compare owing to several confounds that obscure why such differences exist, as well as lacking a ‘visual’ benchmark to contextualise these spatial discrimination abilities. Prior studies examining auditory/haptic approaches (either directly, or indirectly) have varied in the following ways: the type of spatial information being provided (e.g. 2D luminance vs a single-point depth sensor – Bermejo et al., 2015); the spatial resolution being encoded on devices (12 x 12 – Sampaio et al., 2001; 20 x 20 – Nau et al., 2013; 176 x 64 – Striem-Amit et al., 2012; Haigh et al., 2013); variations in the temporal resolution (TDU/BrainPort updating in real-time; the vOICe updating once-per-second); the presentation of information to the user (e.g. ‘all-at-once’ for the TDU/BrainPort or ‘column-by-column’ for the vOICe); there is also an inconsistent use of abstract mappings (e.g. verticality is represented spatially in the BrainPort but via pitch in the vOICe); and finally, without a ‘visual’ condition using this same information to contextualise performance it is unclear how much information is ‘lost’ in the translation and what the upper-bound of performance is likely to be (Cha, Horch, & Normann, 1991). By keeping these aspects consistent across visual, auditory, and haptic conditions, it becomes possible to establish the extent of information loss that occurs from transforming between modalities while enabling a ‘fairer’ comparison.

To address these gaps in knowledge, the present study compares SSDs that output the same spatial resolution (16×8) into spatially distributed patterns of visual, auditory, or tactile stimulation. Users are then evaluated in terms of spatial acuity when using these devices in order to determine the extent of information loss incurred from a purely spatial transformation for sensory substitution. The Synaesthetre device (auditory) and the VibroVision vest (haptic) were selected for comparison as they can output the same spatial resolution into spatialized sound or touch, making them an ideal comparison. In order to

establish the upper limit possible for using a 16×8 depth map, a final condition of viewing this information through spatialized patches of light was created (also using the Synaestheatre software). To examine their spatial acuity, participants were tasked with identifying which of two stimuli was either highest (in the vertical condition) or closest (in the distance condition), using the Synaestheatre (visual), VibroVision (haptic), and Synaestheatre (auditory) in turn. Through a comparison of visual, haptic, and auditory approaches, the spatial information loss attributable to transformations between modalities from using traditional SSD designs can be established.

Hypotheses

Given that the participants were healthy sighted individuals who will be most familiar with using their vision for spatial perception, it is extremely likely that they will be most successful at spatial awareness tasks using their eyesight, compared to haptic or auditory approaches (H_1). It is also expected that participants will yield a greater accuracy on the depth perception tasks compared to verticality (H_2) across all sensory modalities; this is due to the form of the SSDs' output, the signal for depth being **primarily** represented **through** amplitude (brightness, loudness, vibration intensity), compared to verticality which requires the user to perform a more complex interpretation of the signal to discriminate positions (detailed below). As the spatial resolution is identical for both the Synaestheatre and the VibroVision, it is possible that the users will perform equally well on the spatial perception tasks (H_0) indicating that the technology provides the limit on discrimination, rather than the user's discrimination abilities with each modality.

Method

Participants

Right-handed participants ($N = 22$; age = 24 ± 1 years; 16 female, 6 male) were recruited via internal university communications and were predominantly of an academic background; participation was incentivised by the possibility of winning £25, and an opportunity to use novel technology. All the participants read an information sheet about the experiment and sign a written form of consent; Ethical clearance was provided by the host institution. Participants were also instructed that they may withdraw from the experiment at any time of their choosing. Only participants who had normal, or corrected-to-normal, eyesight and who explicitly reported normal hearing were accepted. No participants had prior experience of using sensory substitution devices.

Materials

Synaestheatre

The Synaestheatre device converts visuospatial information into an audio signal using a 3D camera sensor (Structure Sensor, Occipital, USA) mounted on a virtual reality headset (Z4, BOBOVR, China), and a smartphone application (SE, Apple, USA; coded in objective C). The application converts the depth image relayed from the sensor into grey-noise. Grey-noise was chosen as it is perceptually the same loudness across all frequencies, making modulation by HRTF easier to identify. The audio sample is filtered by Panorama (version 5.86, Wavearts, 2016) for Reaper audio software (version 5, Cockos, 2015) which provides HRTF modulation to the grey-noise audio samples. To replicate the HRTF characteristics, the Panorama plug-in filter creates a virtual sound-source and a virtual listener, then synthesises a stereo audio signal to replicate the effect of a sound source 2m away from a human head, varying in azimuth and elevation. In order to align the depth-map and the audio-file HRTF locations, the field-of-view (FoV) of the camera (horizontal = 58° ; vertical = 45°) was separated into a 16×8 resolution, allowing us to generate 128 audio files at combinations of

the following increments; horizontal degrees -29, -25, -21, -17, -14, -10, -6, -2, +2, +6, +10, +14, +17, +21, +25, +29; vertical degrees -23, -16, -10, -3, +3, +10, +16, +23. The measurements to calculate the HRTF were appropriated from a KEMAR head dummy, modelled from the CIPIC HRTF database (Algazi et al., 2001). The sound source was 500ms of grey noise, with HRTF varying its received inter-aural timing and intensity (for horizontal localisation), and spectral frequency content (for vertical localisation).

[please insert figure 1 here]

Grey noise is ideal for using HRTF to interpret verticality as it is only possible to distinguish the vertical locations of ‘broadband sounds’ which contain a wide range of frequencies (e.g., noise, music, or speech). By contrast, it is not possible to tell the vertical location of a single frequency from hearing alone (Blauert, 1997). This is due to the shape of the pinnae (outer ear). When sound bounces off the pinnae, and the source is from a higher vertical place, it preserves more high frequency content, giving the perceived sound a ‘tinnier’ quality. When the sound source is from a lower place, the higher frequencies are suppressed, giving the perceived noise a quality of bass (Parise, Knorre, & Ernst, 2014). Since grey noise is the same loudness across all frequencies, this makes any subtle sound changes caused by vertical HRTF easier to discriminate for the user (see figure 1). Closer distances are represented through increasing loudness to the user. The combination of HRTF and depth-into-loudness allow the user to locate objects in 3D space. During sonification of the depth map, a slight horizontal timing offset is applied to help separate left/right objects; however, changes in signal are instantly responsive to movement of the head or objects. In the visual condition, participants view the Synaestheatre’s depth map directly through a stereo image displayed on an iPhone screen which is mounted in a mobile VR headset (see

figure 2). For the end user wearing the VR headset, this is experienced as viewing a 2D ‘screen’ in front of them with a 16×8 resolution, the horizontal and vertical position of external objects are mapped to their respective horizontal and vertical pixel locations, with the proximity of objects being conveyed through the brightness of the pixels on the screen.

[please insert figure 2 here]

VibroVision

The VibroVision vest is a device developed by Wacker and colleagues (2016) that converts visuospatial information into a two-dimensional vibration image on the abdomen. Like the Synaesthetre device, it converts space into a 16×8 depth map matrix (128 depth points), but instead of the output signal being provided as unique sound files, the information is outputted through 128 eccentric rotating mass pager motors (see figure 3). For the end user the presence of an object picked up by the sensor is felt as a distinct patch of vibration on the abdomen, the horizontal and vertical spatial location of the stimuli in the sensor image is mapped to the horizontal and vertical position on the abdomen, with closer objects making the vibrational patch on the abdomen vibrate more intensely. The spatial information is gathered using a 3D camera (Xtion PRO LIVE, ASUS, Taiwan) mounted to the chest area on the vest; the depth information is relayed to a small computer (2, Raspberry Pi, UK) and then is processed to provide a signal for the vibration motors; processing and vibration motor control is coded in Python, using the OpenCV library. Each electric motor vibrates at a frequency of between 183-233 Hz and is sized at 2.25 cm^2 , with an average centre-to-centre inter-tactor spacing of 2.98 cm horizontally and 3.04 cm vertically. This creates an overall stimulation space of 47cm horizontally by 23.5 cm vertically centred on the subject’s torso.

The centre-to-centre inter-tactor spacing of the VibroVision vest exceeds the 2 cm two-point threshold for the abdomen for the 18-28 age range of our participants (Stevens & Choo, 1996 – also see Lederman & Klatzy, 2009; Weinstein, 1968). While the spatial acuity for vibrotactile stimuli on the abdomen remains an open question (although see Cholewiak, Brill, & Schwab, 2004), the spatial discrimination of vibrotactile stimuli on the back using 4 cm inter-tactor spacings can reach up to 70-77% correct responses in a 3AFC task (Jóhannesson et al., 2017). Since the abdomen has a more sensitive two-point discrimination threshold than the back, the abdomen may also have a higher spatial acuity for vibrotactile stimulation. The vest was calibrated to the same scale depth map as the Synaestheatre, with the FoV of the sensor at 58° (horizontal) and 48° (vertical). For the present study, the vest was powered with mains 240 V electricity through a cable; however, an option exists to use a 5 A LiPo battery pack for extra manoeuvrability.

[please insert figure 3 here]

Both the Synaestheatre and VibroVision were given a maximum sensing range of 2m so that only stimuli were detected by the sensor. Distance was conveyed to users via a linear increase in intensity ramping up from 2m to 0m, manifesting as either light intensity, auditory amplitude, or vibrotactile amplitude. **In addition, as a result of the visual angle between each ‘pixel,’ at certain distances (e.g. closer than 96.4 cm) the stimulus discs could be spread across additional ‘pixels,’ producing a larger area of stimulation (see figure 4).**

[please insert figure 4 here]

Stimuli

Each stimulus consisted of a wooden base (20 cm × 20 cm x 4 cm) with a 1.2 cm hole drilled in the centre. A dowel rod (1.2 cm × 150 cm), marked in 1 cm increments, was inserted in the hole flush with the bottom of the base of the stimuli. The final piece of each stimulus (the head) was a cardboard disc (diameter = 34 cm) with a peg attached to the centre, to allow the circle to be moved to any position on the dowel and, therefore, measure verticality. Two parallel tape measures (200 cm from the participant test mark) were placed on the floor at 100 cm apart to measure depth.

Procedure

Participants were given minimal training and instruction to ensure that results were indicative of device intuition and not meticulous practice. This focus on initial user abilities expands upon our earlier work on the initial impressions of SSDs by potential blind users, who expressed a desire for devices that intuitively conveyed spatial information (Hamilton-Fletcher et al., 2016). A positive initial user experience can bypass the difficulties cited by current SSD experts during their initial learning phase and lower the barrier-to-entry for long-term adoption and accessing the benefits of SSD expertise (Ward & Meijer, 2010).

Participants were shown depth map representations of the disc stimuli (see figure 2) while not blindfolded to aid them in understanding how they would perceive the stimuli once sight was removed. The presentation of stimuli for all practice trials and conditions was fully-randomised in Microsoft Excel as to whether the left or right disc stimulus was the correct answer (i.e. higher, or closer). Before completing the experimental trials, each participant would complete two practice trials using each device, in which they would receive feedback as to whether they made a correct guess. During the experimental trials, the participants received no feedback as to whether they made correct guesses.

The experiment followed a within-subject staircase design, with all participants using every device to complete both depth and verticality spatial perception tasks. All distances and measurements described with relation to the stimuli are using their central points. The depth task involved standing on a mark on the floor in front and centre of the two parallel stimuli (at 100 cm apart); the disc stimuli were set 100 cm from the floor; and the ‘central point’ from which the disc stimulus locations varied was set 120 cm from the participants’ location (see figure 5). The participant was then required to answer the question which stimuli they thought was closer to them (their left, or right), by indicating verbally with ‘left’ or ‘right’. The experimenter writes down each response, which also informs them as to whether the difficulty is increasing or decreasing for the subsequent trial. The distance between the stimuli was initially set at ± 50 cm (e.g. the right stimuli would be set 50 cm closer from the central point, and the left stimuli would be 50 cm further from the central point). This distance would reduce by 15 cm for every correct guess, thereby increasing the difficulty in identifying the difference between the stimuli. When the participant eventually made a mistake, a reversal would occur, increasing the distance by 5 cm. From then on, each correct guess would decrease the difference by 5 cm, and each incorrect guess increases the difference by 5 cm. The stimuli would always increase or decrease relative to the central point; this was done to ensure a uniform increase/decrease in difficulty across all devices and participants. There were ten trials for depth perception for each of the devices. The benefits of following this type of adaptive staircase procedure, as opposed to the standard ‘3-down-1-up’ method (Cornsweet, 1962), is that the number of trials could be decreased drastically. Due to the manual movement of stimuli, each trial was time costly; Therefore, it was practical to reduce the number of trials to minimise the overall experimentation time. However, one potential downside of this method was that if a participant was to make a mistake in the first few trials, and this mistake was not indicative of their actual ability at

using the device, it could be impossible for them to reach their actual performance limit in the remaining trials. This was controlled by allowing subjects to return to increasing the difficulty in increments of 15cm on correct answers, if a mistake was made in one of the first two trials, followed by 3 correct answers in a row. This ensured that a plateau could always be reached by the subject within the ten trials. In the study, this rule ended up only being used for one subject with one SSD. Our pilot testing used 2 subjects to explore a variety of protocols (e.g. performance level at each difficulty; 1-down-1-up only), however, the time taken, and number of trials made alternative approaches impractical. Further testing allowed us to estimate the number of trials at which a participant would hit their peak performance; this typically occurred at trial number six. To further ensure that subjects had reached the limit of their spatial discrimination abilities, each participant's dataset was visually inspected to ensure that their performance had reached a plateau; this was the case for all participants using each sensory modality.

[please insert figure 5 here]

The protocol for testing vertical perception was similar to the protocol used in the depth perception condition. The stimuli were placed along the central point at 120 cm from the participant. The disc stimuli were offset from one another relative to the 'middle point' 100 cm above floor level, starting at +/- 50 cm from this point for each disc stimulus (see figure 6). The measurement for verticality was taken from the centre point of each stimulus disc; therefore, considering the disc's diameter (34 cm), if a participant managed to score an accuracy of 20 cm, each stimulus disc would effectively be overlapping the other by 14 cm in terms of vertical space. Measurements were taken in this way to diminish the time between trials, and to ensure accurate data collection. As with the depth perception, the participant

would answer which one did they perceive as higher (their left, or right), and the distance would reduce by 15 cm until the participant made a mistake. After a reversal, changes to the difference would occur at 5 cm increments. All changes in the distance between the stimuli heads were anchored around the 100 cm mark on the dowel rods. There were ten trials for vertical perception for each device.

[please insert figure 6 here]

The first data collected from each participant was in the control condition. The participant used the Synaestheatre device with the sound turned off, and the visual depth map turned on. This condition provides excellent control data, as both devices were calibrated to the same 16×8 depth map, therefore, in sighted individuals, this represented the highest probable accuracy. The second device that the participants used was the VibroVision vest, during this condition, the participants were blindfolded, and wore sound-reducing earplugs (to minimise the sound effect of the vibrating motors). The vest was tightly wrapped over one layer of participant's clothing, and if the participant was of a slight build, additional cling-film was wrapped around the vest to make certain that the motor-matrix had sufficient contact with the torso. Finally, the participants used the Synaestheatre with the visual depth map switched off, and the sound switched on. During the tasks, subjects stood in a fixed stationary position, from which they could 'look around' using the sensor through varying its pitch, yaw and roll. For every device, the participant would complete the verticality trials first, followed by the depth trials. The visual condition came first as it was the easiest and was useful to familiarize participants with the task. The head-mounted camera used in the visual and auditory conditions caused minor discomfort to wear so was chosen as the first and

last task. These pragmatic considerations come at the expense of counter-balancing which should be considered by future studies. Finally, after each trial, the participant would give a confidence rating of their answer from 1-10 (1 = complete guess, 10 = completely certain).

Analysis

Final scores were taken as the average from the last three trials of each condition and analysed using SPSS (version 23, IBM, 2015). Six dependent variables were created including: visual SSDs for vertical tasks (V-Vert), visual SSDs for depth tasks (V-Depth), haptic SSDs for vertical tasks (H-Vert), haptic SSDs for depth tasks (H-Depth), auditory SSDs for vertical tasks (A-Vert), and auditory SSDs for depth tasks (A-Depth). Any outliers (of which there were two) were detected and removed by using the Outlier Labelling Method with $g = 2.2$; the reason for which was due to a relatively small sample size, and other methods (including the common Outlier Labelling Method, with $g = 1.5$) tending to perform on the conservative side when dealing with smaller groups (Hoaglin, Iglewicz, & Tukey, 1986). The data were checked for normality by examining skewness and kurtosis, in addition to visually inspecting the plots. Furthermore, to check for any violations of sphericity, Mauchly's Test of Sphericity was employed. A one-way repeated measures analysis of variance (ANOVA) was conducted, and subsequent post hoc pairwise comparisons made using the Bonferroni correction method. Effect sizes were calculated using Cohen's d formula by dividing the mean difference by the pooled standard deviation. The participants' confidence scores for each trial were separately compiled with mean scores calculated on a trial-by-trial basis.

Results

After data were imported into SPSS, the Outlier Labelling Method (Hoaglin et al., 1986) established the acceptable bounds for each variable (V-Vert = 0 to 6.34 cm; V-Depth =

0 to 6.34 cm; H-Vert = 0 to 152.33 cm; H-Depth = 0 to 96.34 cm; A-Vert = 0 to 102.99 cm; and A-Depth = 0 to 50.67 cm), and that two outliers existed, which were subsequently removed from the dataset. Visual inspection of the data points on a histogram, in addition to examining skewness and kurtosis for V-Vert (skewness = 1.54, $SE = .51$; kurtosis = **3.99**, $SE = 0.99$); V-Depth (1.33, $SE = 0.51$; and **4.41**, $SE = 0.99$, respectively); H-Vert (skewness = 0.976, $SE = 0.51$; kurtosis = 0.13, $SE = 0.99$); H-Depth (skewness = 1.16, $SE = 0.51$; kurtosis = 1.03, $SE = 0.99$); A-Vert (skewness = 0.86, $SE = 0.52$; kurtosis = 0.01, $SE = 0.99$); and A-Depth (skewness = 0.57, $SE = 0.51$; kurtosis = 1.24, $SE = 0.99$), revealed that the visual data was not normally distributed (see bolded). However, one-way repeated measures ANOVAs are known to be quite robust to violations of normal distribution (Field, 2013, p. 444), with the F -statistic relatively unaffected by variations of normality when condition sizes are equal (Donaldson, 1968), which is the case here, making transformations unnecessary (Glass, Peckham & Sanders, 1972). Therefore, analysis was continued (see table 1 for variable mean scores, visual angle, standard deviation, and confidence intervals).

Table 1: Mean spatial discrimination scores, visual angle, standard deviations and confidence intervals for the visual, haptic, and auditory devices.

Condition	SSDs	Mean (cm)	Visual angle	S.D. (cm)	Lower Bound	Upper Bound
Verticality	Vision	2.09	0.99°	1.24	1.50	2.66
	Haptic	44.42	20.97°	29.90	30.42	58.41
	Auditory	30.08	14.29°	18.88	21.25	38.92
Depth	Vision	2.34	<i>n/a</i>	1.16	1.79	2.87
	Haptic	28.83	<i>n/a</i>	20.59	19.20	38.47
	Auditory	8.25	<i>n/a</i>	7.26	4.85	11.65

Mauchly's Test of Sphericity showed that the assumption of sphericity was not met for verticality ($\chi^2(2) = 6.43$, $p = .04$) or depth ($\chi^2(2) = 19.97$, $p < .001$), therefore Huynh-Feldt

and Greenhouse-Geisser estimates of sphericity correction were applied respectively ($\varepsilon = .77$ and $\varepsilon = .60$). The repeated measures ANOVA suggested that there were main effects for both the verticality ($F(1.65, 31.29) = 24.44, p < .001, \eta^2 = .56$), and depth tasks ($F(1.20, 22.75) = 26.19, p < .001, \eta^2 = .58$), therefore post hoc pairwise comparisons were made using a Bonferroni adjustment (see table 2 for mean differences for verticality, and table 3 for mean differences for depth).

Table 2: Mean difference between devices for verticality.

Mean Difference ($X_i - X_j$) (cm)			
	Vision	Haptic	Auditory
Vision	-	42.33**	28.00**
Haptic	42.33**	-	14.33
Auditory	28.00**	14.33	-

Note: ** = $p \leq .001$; * = $p \leq .05$. Repeated values in grey.

For discriminating verticality, users were significantly more accurate with visual SSDs than both haptic ($p < .001, d = 2.65$) and auditory SSDs ($p < .001, d = 2.67$); there was no significant difference in performance between haptic and auditory SSDs.

Table 3: Mean differences between devices for depth.

Mean Difference ($X_i - X_j$) (cm)			
	Vision	Haptic	Auditory
Vision	-	26.50**	5.92*
Haptic	26.50**	-	20.58**
Auditory	5.92*	20.58**	-

Note: ** = $p \leq .001$; * = $p \leq .05$. Repeated values in grey.

For discriminating depth, users were significantly more accurate with visual SSDs than both haptic and auditory SSDs ($p < .001, d = 2.44$; and $p = .01, d = 1.41$ respectively).

However, unlike the outcomes for the verticality tasks, in the depth condition, users were significantly more accurate with auditory SSDs than haptic SSDs ($p = .001$, $d = 1.48$).

In order to compare whether discriminating verticality or depth differed in their difficulty, a series of Bonferroni-corrected paired sample t-tests were performed with each device. On average, with the visual SSD, subjects performed slightly better with vertical discrimination ($M = 2.09$, $SE = 0.28$) than depth discrimination ($M = 2.34$, $SE = 0.26$), this difference, of -0.25cm, was not significant, $t(19) = -0.59$, $p = .561$, $d = 0.21$. For the haptic SSD, subjects performed slightly worse with vertical discrimination ($M = 44.42$, $SE = 6.69$) than depth discrimination ($M = 28.83$, $SE = 4.60$), this difference, of 15.58cm, was not significant, $t(19) = 1.76$, $p = .095$, $d = 0.61$. For the auditory SSD, subjects discriminated verticality to an average of 30.08cm ($SE = 4.22$), with substantially superior depth discrimination abilities of 8.25cm ($SE = 1.62$), this difference of 21.83cm, was highly significant, $t(19) = 5.94$, $p < .001$, $d = 1.53$. Overall this indicates that discriminating verticality and depth was of a similar difficulty for visual and haptic approaches, but that depth was significantly easier to discriminate than verticality for auditory SSDs.

For confidence ratings, vision had the highest mean across all ten trials for vertical perception (see figure 7, upper row), and had a total combined average of 7.34 ± 1.67 . The mean scores for haptic and auditory were very similar at 4.32 ± 0.63 , and 4.26 ± 0.58 , respectively. Regarding depth perception (see figure 7, lower row), vision had a higher total mean score than both haptic and auditory, scoring at 7.24 ± 2.01 , compared to 5.59 ± 0.56 (haptic), and 5.44 ± 0.97 (auditory). However, examination of figure 7 shows that visual, despite initially holding a high level of confidence rating on the first six trials, rapidly declines to below haptic at trial number ten (4.00 ± 2.32 verses 4.60 ± 2.13) and only slightly holds above auditory (3.95 ± 1.91).

[please insert figure 7 here]

Discussion

The present research aimed to compare spatial perception by users when the same information (a 16×8 depth map) is perceived as either spatialized light, sound, or touch using sensory substitution. This approach allows a quantification of the level of information loss attributable to transforming between modalities in traditional SSD designs across two tasks (vertical and depth localisation) as well as their confidence in making these assessments. In the verticality condition, visual SSDs had a significantly higher accuracy than our auditory and haptic approaches, discriminating between visual angles of 1° , 14° , and 21° respectively. For depth perception, visual feedback was once again significantly more accurate than auditory and haptic feedback in our SSDs, however this time auditory feedback significantly outperformed haptic feedback, with discrimination abilities at 2cm, 8cm, and 29cm, **for visual, auditory, and haptic SSDs respectively**. Users had the highest levels of confidence using visual feedback; however, despite users being equally confident in using haptic and auditory SSDs, auditory specialisation appeared to yield higher accuracies relative to haptic approaches.

For both vertical and depth perception tasks, the visual SSDs were significantly more accurate than SSDs utilising alternative sensory modalities (H_1 was accepted). As predicted, our participant groups had superior spatial perception from using visual information. However, their degree of spatial acuity was more surprising, for objects at 1.2m, users could interpret visual information to within just a few centimetres (V-Vert = 2.09 ± 1.24 cm; V-Depth = 2.34 ± 1.16 cm). Considering that all the devices work from identical 16×8 depth maps, the visual scores likely represent the best possible performance from this information, providing useful knowledge about the practical capacity of using ‘low’ spatial resolutions in

SSDs, and create a benchmark to which users of audio and haptic SSDs could conceivably improve to with additional training and/or design changes.

These ‘visual-into-visual’ SSDs may allow experimenters and designers to quickly assess the effectiveness of the information provided and establish the upper-bound of performance users can get from this information. This provides the advantage of eliminating the additional confounds that arise from converting between modalities (Cha et al., 1991), while also reducing the time required to train users on how to interpret auditory or tactile information in a visual manner, or reach ‘expertise’ (Ward & Meijer, 2010; Ortiz et al., 2011). This considered, care must also be taken to represent the final desired output fairly in vision, for example, the vOICe SSD outputs information serially, column-by-column to the end user, and as such, any visual comparison should also output information in a similar manner such as through the use of column-like apertures (Morgan, Findlay, & Watt, 1982; Day & Duffy, 1988; Mateef, Popov, & Hohnsbein, 1993; James, Huh, & Kim, 2010). These parallels also provide new opportunities for SSD design to learn from how this information can be more efficiently delivered and processed by the end user (Króliczak, Goodale, & Humphrey, 2003; Craddock, Martinovic, & Lawson, 2011).

Comparing the difficulties of the vertical discrimination task and the depth discrimination task revealed that while for the visual and haptic SSDs these tasks did not significantly differ in their difficulty, for the auditory SSDs discriminating verticality was significantly more difficult than discriminating depth (H_2 was partially-accepted). This is likely due to the differences in difficulty between discriminating two bursts of noise that varied in loudness (as required in the depth task) and discriminating between two bursts of noise that varied in spectral frequency content (as required in the verticality task). The lack of familiarity with vertical hearing, or the complexity in terms of discriminating between subtle changes in frequency are likely to underscore these differences. **It could also be stated that**

the depth task also provided additional redundant information in terms of the total ‘area of stimulation’ and that it may be unclear whether subjects were utilising intensity or this. However, the discrimination thresholds reached by subjects in all SSDs surpassed the distance at which stimuli took up additional ‘pixels,’ making this cue unavailable. This could be further evaluated in the future through utilising stimuli of different physical sizes.

These results provide an important baseline for SSDs, in assessing translation methods that piggyback off of veridical spatial perception for each sense. In turn, these baselines can inform future SSD design by comparing changes in design with changes in performance. To date, many different types of SSD have been reported in the scientific literature, utilising a huge variety of signal outputs (Wacker et al., 2016; Ertan et al., 1998; Rochlis, 1998; Jones et al., 2004; Spanlang et al., 2010; Ward & Meijer, 2010; Hamilton-Fletcher & Ward, 2013; Hamilton-Fletcher et al., 2016; Jóhannesson et al., 2016). Historically, these devices have been difficult to compare against each other, across different tasks, or against a reasonable benchmark, limiting our knowledge on effective approaches to turning vision into sound or touch. We propose that a comparison of the same information – either kept within the source modality (e.g. vision) or transferred to another (e.g. hearing, touch), allows the most accurate assessment of how well the translation to other modalities works within these devices.

With a baseline comparison of ‘visual-into-visual’ SSDs in place, any changes in sensory substitution design that reduce any ‘modality gap’ between visual and audio/haptic approaches can be evaluated. Knowing the upper-limit of performance on an SSD can set expectations in how best to present this information to the user. This covers a huge range of design considerations, from utilising the substituting senses’ ability to discriminate and categorise information (e.g. utilising the principles of spatialized hearing - Blauert, 1997, or how ‘auditory objects’ are perceived from an audio stream - Bregman, 1994), a consideration

of users' differing abilities (e.g. early-blind have impaired vertical hearing localisation – Zwiers, Van Opstal, & Cruysberg, 2001; Lewald, 2002; but superior tactile acuity, discrimination of auditory pitch, loudness, and horizontal localisation cues – Goldreich & Kanics, 2003; Norman & Bartholomew, 2011; Wan et al., 2010; Röder et al., 1999; Fieger et al., 2006; Gougoux et al., 2004; Kolarik, Cirstea, & Pardhan, 2013), and when appropriate, taking advantage of the users' multisensory processing biases (e.g. cross-modal correspondences – Hamilton-Fletcher, Wright, & Ward, 2016; Hamilton-Fletcher et al., 2017; 2018; Deroy et al., 2016). Beyond this, many other factors remain to assess, such as: selecting task-appropriate information; appropriate spatial, temporal, and colour resolutions; avoiding sensory and attentional overloading; training and usability in daily life; and finally, changes in perception, externalisation, and qualia (Hamilton-Fletcher & Ward, 2013; Ward & Meijer, 2010; Kristjánsson et al., 2017; Ortiz et al., 2011; Bertram & Stafford, 2016; Brown, Simpson, & Proulx, 2014; 2015; Brown & Proulx, 2016; Auvray, Hanneton, & O'Regan, 2007; Hartcher-O'Brien & Auvray, 2014).

When comparing between the substituted modalities (hearing, touch), there was partial support for H_1 for the difference between haptic and auditory devices, specifically for the assessment of depth, with the Synaestheatre being significantly more accurate than the VibroVision vest (by 20.58 cm). **This suggests that comparing two objects that differ in amplitude (and sometimes area of stimulation), is easier to discriminate using our auditory rather than our tactile output.** While vertical localisation **performance was superior** for the Synaestheatre (14°) than the VibroVision (21°), this is not to a significant degree. In terms of vertical localisation within hearing, users are typically very limited (3.65° – Perrott & Saberi, 1990; with movement increasing this resolution – Strybel, Manligas & Perrott, 1992). As such, there is room to improve spatial localisation in audio SSDs to reach this level, most likely through personalised rather than generic HRTFs (reaching 9.47° – Pec et al., 2008).

There also exists the possibility of exaggerating natural elevation cues from the pinnae or spatial coordinates for peripheral stimuli, although, it is currently unclear if any perceptual or spatial exaggerations can be effectively utilised or integrated with veridical hearing. The most common alternative is the use of cross-modal correspondences, such as substituting height for pitch, which may create easier to discriminate signal changes for users (Striem-Amit et al., 2012; Haigh et al., 2013), however this can have similar limitations (Brown et al., 2014; 2015; Brown & Proulx, 2016), and while this also has the advantage of being intuitive for those with prior visual experience, it is potentially not intuitive for the congenitally blind (Eitan, Ornoy, & Granot, 2012; Deroy et al., 2016). In addition, the use of pitch for height limits the use of pitch for other qualitative content such as texture, colour, or shape (Hamilton-Fletcher et al., 2016; 2017; 2018).

The tactile approach taken by the VibroVision resulted in errors in vertical localisation at 44.42cm (± 29.90), and depth localisation at 28.83cm (± 20.59) for objects at 1.2m. Tactile approaches also showed the largest amount of variation in user abilities (see table 1). Some of this may be due to physical factors limiting skin resolution, such as variations in clothing insulation or body size influencing the location of the vibrotactile pads. Jóhannesson et al. (2017) examined the vibrotactile spatial acuity of the back, providing important implications for vibrotactile SSDs. They note that as the distance between tactor motors is reduced, while total information increases, user accuracy decreases. Furthermore, they found that tactors mounted to sponge with a 3 cm distance (centre to centre) were significantly more accurate than tactors mounted to fabric (as with the VibroVision) at 4 cm distance (centre to centre). They suggest that this is due to the sponge aiding in reducing the amount of vibration distribution from the tactors, allowing a more localised sense of vibration. Van Erp (2005) attached tactor motors directly to the skin, not practical for

widespread use, but in the case of a single individual using this technology, direct skin-tactor contact provided improved accuracy.

Within the present study, subjects were able to conduct a limited range of movement with the sensor, and previous studies have shown that this form of self-initiated motion can also improve users' spatial abilities, for example, when viewing an object, manipulating the position of the sensor will produce predictable movement of the object's tactile representation on the body, allowing users to both find the threshold at which an object's signal moves to a new tactor, as well as easily predict which tactor is most likely to become active based on this movement. This predictability in signal change reduces the ambiguity of which tactors are active, effectively increasing the users' functional resolution (Van Erp, 2005). In discriminating multiple tactors, the role of training is particularly important, with even short training sessions improving spatial discrimination by 36% and intensity discrimination by 44% (Stronks et al., 2017). These considerations could help narrow the gap between the VibroVision and alternative modalities while keeping the advantages inherent to the tactile channel, such as no auditory inference and using areas of the skin not essential for daily life.

Outside of SSD research, previous attempts to compare the processing of similar spatial information through vision or touch (via the fingertips) have reduced the visual resolution of the stimuli in order to simulate equivalent receptor densities between modalities. When this kind of equivalent information is presented via image blurring, comparable performance has been found for letter and Braille stimuli presented as blurred images or raised patterns (Loomis, 1981; 1982), however further differences have emerged for how the modalities process dot and joined-Braille stimuli, with touch better at separating out line **distractors and vision better at separating out** dot distractors (Loomis, 1993). When equivalence is reached through miniaturizing the visual image, functional equivalence was

reached for images sized at 0.037° visual angle and 5 mm for passive tactile stimulation (4 mm for active exploration), which corresponds to similar multiples of receptor spacing - whether SA for touch, or cones for vision (Phillips, Johnson & Browne, 1983). Cho et al (2015) compared spatial patterns on larger stimuli that spanned either ~ 100 SA1 afferents (1 cm^2) through tactile stimulation or ~ 100 retinal cones (0.083° visual angle) via visual stimulation and found superior performance for visual presentations. This suggests that visual performance increases more than tactile when increasing numbers of afferents are involved. Relevant to the present SSD research, while subjects can discriminate vibrational stimuli at closer distances than the inter-tactor spacings used by the VibroVision vest, this refers to minimum thresholds for discrimination, not perfect performance (Jóhannesson et al., 2017), which means that the ability to resolve points of stimulation for the haptic SSD is likely inferior than for visual SSDs.

We observed an interesting disparity between participant confidence and abilities. Predictably, confidence was highest with vision, until the task became vastly more difficult in the last few trials, here participants reduced in confidence, but not ability. For haptic and auditory SSDs, confidence was similar throughout the trials, despite differences in ability. For depth, the auditory Synaesthete was significantly more accurate than the VibroVision vest (by 20.58 cm), and although statistically non-significant, again auditory perception was more accurate than haptic for discriminating verticality (by 14.33 cm). This discrepancy may occur from users overestimating their tactile localisation/intensity discrimination abilities, or tactile sensations may be perceived as more intuitive for users to interpret from. It is unclear to what extent this lowering of confidence from visual SSDs to auditory or haptic versions is due to the increased difficulty with discriminating the outputted signal, or unfamiliarity with utilising auditory/haptic information in this way. One way to disentangle these competing explanations is by introducing signals that are easier to discriminate (e.g. verticality

represented by pitch-changes, or tactile-codes), to see if these can reach parity with visual SSDs in confidence as well as ability. Future studies may also wish to examine the difference in ability and confidence of blind end users in using SSDs to solve spatial tasks, especially since blind individuals are likely to focus more extensively on hearing and touch to perform spatial discrimination tasks in daily life. In terms of SSD training, users may feel more inclined to continue to learn and practice with SSDs they are confident using; nevertheless, if a user's confidence and actual abilities are mismatched this could have negative consequences. Understanding how users evaluate their abilities (metacognition) will be essential to furthering training regimes and improving user abilities and safety.

Sensory substitution devices that provide a sense of spatial perception have the capacity to alter both the perceptions and representations of space in users with visual impairments or blindness. As previously noted, individuals with no visual experience tend to have more separated spatial reference frames, prefer egocentric co-ordinates for representing space, have impaired spatial memory/recognition, less flexibility in navigation, distorted representations of space, impaired vertical hearing, and find it harder to replicate spatial relationships relative to those with prior visual experience such as the sighted, or late-blind (Pasqualotto & Proulx, 2012; Iachini et al., 2014; Corazzini et al., 2010; Ruggiero et al., 2018; Vercillo et al., 2018; Hötting & Röder, 2004; Hötting et al., 2004; Lewald, 2002; Pasqualotto et al., 2013; Kolarik et al., 2017; Gori et al., 2017). Despite any differences, functional equivalence is also frequently observed on a variety of spatial tasks irrespective of the participant's experience of sight or their given information source, whether from vision, hearing, touch, or language (for a review see Giudice, 2018). Any differences in how various sightedness groups tend to discriminate, represent, and utilise the available information need not be inevitable or immutable, rather, that additional training and information can change all of these factors. Their use of sensory substitution devices can reveal what specific spatial

information is key to integrating, calibrating, and utilising more effective representations of external space. This has begun to be explored through both fostering specific audio-motor contingencies (Cappagli et al., 2017), however, only a few alterations to spatial perception have been explored with the more ‘visual-like’ SSDs, such as the use of ego/allocentric representations (Pasqualotto, & Esenkaya, 2016). By contrast, having prior visual experience can also carry over vision-specific distortions such as the Ponzo illusion or vertical-horizontal illusion to stimuli sonified by auditory SSDs that the congenitally blind do not experience, allowing a more accurate reconstruction (Renier et al., 2005; Renier, Bruyer & De Volder, 2006). Overall, the reported range of effects to spatial perception found in congenital blindness remains an open area for further exploration for SSDs as an intervention going forward.

While this study examined SSDs that convert visuospatial information into spatialized light, sound, or touch in isolation, more open questions remain as to the ability of users to integrate this spatial information across modalities (i.e. visuospatial into both audio and tactile stimulations). It is currently unclear whether these hybrid SSDs confer any advantages in performance or whether this information can be effectively integrated. An additional question relates to whether distinct information presented to each sense (e.g. visuospatial into touch but colour information into sound) can also be effectively integrated into a unified coherent representation of external space. Part of the promise of hybrid SSDs is the potential to expand past the processing limitations inherent to individual modalities, however this remains to be realised.

Overall, the present study provides evidence of the functional spatial resolutions possible for participants using visual, auditory, or haptic SSDs when the visuospatial information provided remains constant. Our methodology showcases how the ‘upper-limit’ of performance for SSDs can be quickly established and isolated from any reductions in the

functional spatial resolution attributable to converting spatial signals into other modalities using conventional SSD designs. Furthermore, we identify multiple ways in which any reductions in spatial resolution attributable to this transformation can be reduced for a wide range of users. This information can help identify optimal ways of providing visuospatial information to the visually-impaired to further enhance their representation of external space.

References

- Algazi, V. R., Avendano, C., & Duda, R. O. (2001). Elevation localization and head-related transfer function analysis at low frequencies. *The Journal of the Acoustical Society of America*, 109(3), 1110-1122. doi: 10.1121/1.1349185
- Algazi, V. R., Duda, R. O., Thompson, D. M., & Avendano, C. (2001). The cipic hrtf database. In *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE*. doi: 10.1109/ASPAA.2001.969552
- Auvray, M., Hanneton, S., & O'Regan, J. K. (2007). Learning to perceive with a visuo—auditory substitution system: localisation and object recognition with 'The vOICe'. *Perception*, 36(3), 416-430. doi: 10.1068/p5631
- Bach-y-Rita, P., Collins, C. C., Saunders, F. A., White, B., & Scadden, L. (1969). Vision substitution by tactile image projection. *Nature*, 221(5184), 963-964. doi: 10.1038/221963a0
- Bermejo, F., Di Paolo, E. A., Hüg, M. X., & Arias, C. (2015). Sensorimotor strategies for recognizing geometrical shapes: a comparative study with different sensory substitution devices. *Frontiers in Psychology*, 6, 679. doi: 10.3389/fpsyg.2015.00679
- Bertram, C., & Stafford, T. (2016). Improving training for sensory augmentation using the science of expertise. *Neuroscience & Biobehavioral Reviews*, 68, 234-244. doi: 10.1016/j.neubiorev.2016.05.026
- Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT press.
- Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT press.
- Brown, D. J., Simpson, A. J., & Proulx, M. J. (2014). Visual objects in the auditory system in sensory substitution: how much information do we need?. *Multisensory Research*, 27(5-6), 337-357. doi: 10.1163/22134808-00002462
- Brown, D. J., Simpson, A. J., & Proulx, M. J. (2015). Auditory scene analysis and sonified visual images. Does consonance negatively impact on object formation when using complex sonified stimuli?. *Frontiers in Psychology*, 6. doi: 10.3389/fpsyg.2015.01522
- Brown, D. J., & Proulx, M. J. (2016). Audio–Vision Substitution for Blind Individuals: Addressing Human Information Processing Capacity Limitations. *IEEE Journal of Selected Topics in Signal Processing*, 10(5), 924-931. doi: 10.1109/JSTSP.2016.2543678
- Caraiman, S., Morar, A., Owczarek, M., Burlacu, A., Rzeszotarski, D., Botezatu, N., Herghelegiu, P., Moldoveanu, F., Strumillo, P., & Moldoveanu, A. (2017, October). Computer Vision for the Visually Impaired: The Sound of Vision System. In *Computer Vision Workshop (ICCVW), 2017 IEEE International Conference on* (pp. 1480-1489). IEEE. doi: 10.1109/ICCVW.2017.175
- Cappagli, G., Cocchi, E., & Gori, M. (2017). Auditory and proprioceptive spatial impairments in blind children and adults. *Developmental Science*, 20(3), e12374. doi: 10.1111/desc.12374
- Cappagli, G., Finocchietti, S., Baud-Bovy, G., Cocchi, E., & Gori, M. (2017). Multisensory rehabilitation training improves spatial perception in totally but not partially visually deprived children. *Frontiers in Integrative Neuroscience*, 11, 29. doi: 10.3389/fnint.2017.00029
- Cha, K., Horch, K. W., & Normann, R. A. (1992). Mobility performance with a pixelized vision system. *Vision Research*, 32(7), 1367-1372. doi: 10.1016/0042-6989(92)90229-C

- Cho, Y., Craig, J. C., Hsiao, S. S., & Bensmaia, S. J. (2015). Vision is superior to touch in shape perception even with equivalent peripheral input. *Journal of Neurophysiology*, 115(1), 92-99. doi: 10.1152/jn.00654.2015
- Cholewiak, R. W., Brill, J. C., & Schwab, A. (2004). Vibrotactile localization on the abdomen: Effects of place and space. *Attention, Perception, & Psychophysics*, 66(6), 970-987. doi: 10.3758/BF03194989
- Corazzini, L., Tinti, C., Schmidt, S., Mirandola, C., & Cornoldi, C. (2010). Developing spatial knowledge in the absence of vision: allocentric and egocentric representations generated by blind people when supported by auditory cues. *Psychologica Belgica*, 50(3-4). doi: 10.5334/pb-50-3-4-327
- Cornsweet, T. N. (1962). The staircase-method in psychophysics. *The American Journal of Psychology*, 75(3), 485-491. doi: 10.2307/1419876
- Craddock, M., Martinovic, J., & Lawson, R. (2011). An advantage for active versus passive aperture-viewing in visual object recognition. *Perception*, 40(10), 1154-1163. doi: 10.1068/p6974
- Day, R. H., & Duffy, F. M. (1988). Illusions of time and extent when the Müller-Lyer figure moves in an aperture. *Perception & Psychophysics*, 44(3), 205-210. doi: 10.3758/BF03206289
- Deroy, O., Fasiello, I., Hayward, V., & Auvray, M. (2016). Differentiated audio-tactile correspondences in sighted and blind individuals. *Journal of Experimental Psychology: Human Perception and Performance*, 42(8), 1204. doi: 10.1037/xhp0000152
- Donaldson, T. S. (1968). Robustness of the F-test to errors of both kinds and the correlation between the numerator and denominator of the F-ratio. *Journal of the American Statistical Association*, 63(322), 660-676. doi: 10.2307/2284037
- Dunai, L., Peris-Fajarnés, G., Lluna, E., & Defez, B. (2013). Sensory navigation device for blind people. *The Journal of Navigation*, 66(3), 349-362. doi: 10.1017/S0373463312000574P
- Eckert, M., Blex, M., & Friedrich, C. M. (2018). Object detection featuring 3D audio localization for Microsoft HoloLens. In *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies* (pp. 555-561). doi: 10.5220/0006655605550561
- Eitan, Z., Ornoy, E., & Granot, R. Y. (2012). Listening in the dark: Congenital and early blindness and cross-domain mappings in music. *Psychomusicology: Music, Mind, and Brain*, 22(1), 33. doi: 10.1037/a0028939
- Elli, G. V., Benetti, S., & Collignon, O. (2014). Is there a future for sensory substitution outside academic laboratories?. *Multisensory Research*, 27(5-6), 271-291. doi: 10.1163/22134808-00002460
- Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences*, 12(10), 388-396. doi: 10.1016/j.tics.2008.07.004
- Ertan, S., Lee, C., Willets, A., Tan, H., & Pentland, A. (1998, October). A wearable haptic navigation guidance system. In *Wearable Computers, 1998. Digest of Papers. Second International Symposium on* (pp. 164-165). IEEE. doi: 10.1109/ISWC.1998.729547
- Fieger, A., Röder, B., Teder-Sälejärvi, W., Hillyard, S. A., & Neville, H. J. (2006). Auditory spatial tuning in late-onset blindness in humans. *Journal of Cognitive Neuroscience*, 18(2), 149-157. doi: 10.1162/jocn.2006.18.2.149
- Field, A. (2013) *Discovering Statistics Using IBM SPSS Statistics*. London. SAGE Publications Ltd.

- Fristot, V., Boucheteil, J., Granjon, L., Pellerin, D., & Alleysson, D. (2012, August). Depth—Melody substitution. In *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European* (pp. 1990-1994). IEEE.
- Giudice, N. (2018). Navigating without Vision: Principles of Blind Spatial Cognition. In D. Montello (Ed.), *Handbook of Behavioral and Cognitive Geography* (pp. 260-288). Northampton, MA: Edward Elgar Publishing.
- Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, 42(3), 237-288. doi: 10.3102/00346543042003237
- Goldreich, D., & Kanics, I. M. (2003). Tactile acuity is enhanced in blindness. *Journal of Neuroscience*, 23(8), 3439-3445. doi: 10.1523/JNEUROSCI.23-08-03439.2003
- González-Mora, J. L., Rodríguez-Hernandez, A., Rodríguez-Ramos, L. F., Díaz-Saco, L., & Sosa, N. (1999, June). Development of a new space perception system for blind people, based on the creation of a virtual acoustic space. In *International Work-Conference on Artificial Neural Networks* (pp. 321-330). Springer, Berlin, Heidelberg. doi: 10.1007/BFb0100499
- González-Mora, J. L., Rodríguez-Hernandez, A., Burunat, E., Martin, F., & Castellano, M. A. (2006). Seeing the world by hearing: Virtual Acoustic Space (VAS) a new space perception system for blind people. In *Information and Communication Technologies, 2006. ICTTA'06. 2nd* (Vol. 1, pp. 837-842). IEEE. doi: 10.1109/ICTTA.2006.1684482
- Gougoux, F., Lepore, F., Lassonde, M., Voss, P., Zatorre, R. J., & Belin, P. (2004). Neuropsychology: pitch discrimination in the early blind. *Nature*, 430(6997), 309-309. doi: 10.1038/430309a
- Gori, M., Sandini, G., Martinoli, C., & Burr, D. C. (2013). Impairment of auditory spatial localization in congenitally blind human subjects. *Brain*, 137(1), 288-293. doi: 10.1093/brain/awt311
- Gori, M., Cappagli, G., Tonelli, A., Baud-Bovy, G., & Finocchietti, S. (2016). Devices for visually impaired people: High technological devices with low user acceptance and no adaptability for children. *Neuroscience & Biobehavioral Reviews*, 69, 79-88. doi: 10.1016/j.neubiorev.2016.06.043
- Gori, M., Cappagli, G., Baud-Bovy, G., & Finocchietti, S. (2017). Shape perception and navigation in blind adults. *Frontiers in Psychology*, 8. doi: 10.3389/fpsyg.2017.00010
- Haigh, A., Brown, D. J., Meijer, P., & Proulx, M. J. (2013). How well do you see what you hear? The acuity of visual-to-auditory sensory substitution. *Frontiers in Psychology*, 4. doi: 10.3389/fpsyg.2013.00330
- Hamilton-Fletcher, G., & Ward, J. (2013). Representing colour through hearing and touch in sensory substitution devices. *Multisensory Research*, 26(6), 503-532. doi: 10.1163/22134808-00002434
- Hamilton-Fletcher, G., Obrist, M., Watten, P., Mengucci, M., & Ward, J. (2016). I Always Wanted to See the Night Sky: Blind User Preferences for Sensory Substitution Devices. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2162-2174. ACM. doi: 10.1145/2858036.2858241
- Hamilton-Fletcher, G., Pisanski, K., Reby, D., Stefańczyk, M., Ward, J., & Sorokowska, A. (2018). The role of visual experience in the emergence of cross-modal correspondences. *Cognition*, 175, 114-121. doi: 10.1016/j.cognition.2018.02.023

- Hamilton-Fletcher, G., Witzel, C., Reby, D., & Ward, J. (2017). Sound Properties associated with Equiluminant Colours. *Multisensory Research*, 30, 337-362. doi: 10.1163/22134808-00002567
- Hamilton-Fletcher, G., Wright, T. D., & Ward, J. (2016). Cross-modal correspondences enhance performance on a colour-to-sound sensory substitution device. *Multisensory Research*, 29(4-5), 337-363. doi: 10.1163/22134808-00002519
- Hartcher-O'Brien, J., & Auvray, M. (2014). The process of distal attribution illuminated through studies of sensory substitution. *Multisensory Research*, 27(5-6), 421-441. doi: 10.1163/22134808-00002456
- Heller, M. A., & Kennedy, J. M. (1990). Perspective taking, pictures, and the blind. *Perception & Psychophysics*, 48(5), 459-466. doi: 10.3758/BF03211590
- Hoaglin, D. C., Iglewicz, B., & Tukey, J. W. (1986). Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 81(396), 991-999. doi: 10.2307/2289073
- Hötting, K., & Röder, B. (2004). Hearing cheats touch, but less in congenitally blind than in sighted individuals. *Psychological Science*, 15(1), 60-64. doi: 10.1111/j.0963-7214.2004.01501010.x
- Hötting, K., Rösler, F., & Röder, B. (2004). Altered auditory-tactile interactions in congenitally blind humans: an event-related potential study. *Experimental Brain Research*, 159(3), 370-381. doi: 10.1007/s00221-004-1965-3
- Iachini, T., Ruggiero, G., & Ruotolo, F. (2014). Does blindness affect egocentric and allocentric frames of reference in small and large scale spaces?. *Behavioural Brain Research*, 273, 73-81. doi: 10.1016/j.bbr.2014.07.032
- James, T. W., Huh, E., & Kim, S. (2010). Temporal and spatial integration of face, object, and scene features in occipito-temporal cortex. *Brain and Cognition*, 74(2), 112-122. doi: 10.1016/j.bandc.2010.07.007
- Jóhannesson, Ó. I., Balan, O., Unnthorsson, R., Moldoveanu, A., & Kristjánsson, Á. (2016). The sound of vision project: on the feasibility of an audio-haptic representation of the environment, for the visually impaired. *Brain Sciences*, 6(3), 20. doi: 10.3390/brainsci6030020
- Jóhannesson, Ó. I., Hoffmann, R., Valgeirsdóttir, V. V., Unnþórsson, R., Moldoveanu, A., & Kristjánsson, Á. (2017). Relative vibrotactile spatial acuity of the torso. *Experimental Brain Research*, 1-11. doi: 10.1007/s00221-017-5073-6
- Jones, L. A., Nakamura, M., & Lockyer, B. (2004, March). Development of a tactile vest. In *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2004. HAPTICS'04. Proceedings. 12th International Symposium on*(pp. 82-89). IEEE. doi: 10.1109/HAPTIC.2004.1287181
- Kelly, J. W., McNamara, T. P., Bodenheimer, B., Carr, T. H., & Rieser, J. J. (2008). The shape of human navigation: How environmental geometry is used in maintenance of spatial orientation. *Cognition*, 109(2), 281-286. doi: 10.1016/j.cognition.2008.09.001
- Kistler, D. J., & Wightman, F. L. (1992). A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *The Journal of the Acoustical Society of America*, 91(3), 1637-1647. doi: 10.1121/1.402444
- Kolarik, A. J., Cirstea, S., & Pardhan, S. (2013). Evidence for enhanced discrimination of virtual auditory distance among blind listeners using level and direct-to-reverberant cues. *Experimental Brain Research*, 224(4), 623-633. doi: 10.1007/s00221-012-3340-0

- Kolarik, A. J., Pardhan, S., Cirstea, S., & Moore, B. C. (2017). Auditory spatial representations of the world are compressed in blind humans. *Experimental Brain Research*, 235(2), 597-606. doi: 10.1007/s00221-016-4823-1
- Kristjánsson, Á., Moldoveanu, A., Jóhannesson, Ó. I., Balan, O., Spagnol, S., Valgeirsdóttir, V. V., & Unnthorsson, R. (2016). Designing sensory-substitution devices: Principles, pitfalls and potential. *Restorative Neurology and Neuroscience*, 34(5), 769-787. doi: 10.3233/RNN-160647
- Króliczak, G., Goodale, M. A., & Humphrey, G. K. (2003). The effects of different aperture-viewing conditions on the recognition of novel objects. *Perception*, 32(10), 1169-1179. doi: 10.1068/p3443
- Kulkarni, A., Isabelle, S. K., & Colburn, H. S. (1999). Sensitivity of human subjects to head-related transfer-function phase spectra. *The Journal of the Acoustical Society of America*, 105(5), 2821-2840. doi: 10.1121/1.426898
- Lederman, S. J., & Klatzky, R. L. (2009). Haptic perception: A tutorial. *Attention, Perception, & Psychophysics*, 71(7), 1439-1459. doi: 10.3758/APP.71.7.1439
- Lewald, J. (2002). Vertical sound localization in blind humans. *Neuropsychologia*, 40(12), 1868-1872. doi: 10.1016/S0028-3932(02)00071-4
- Loomis, J. M. (1981). On the tangibility of letters and braille. *Perception & Psychophysics*, 29(1), 37-46. doi: 10.3758/BF03198838
- Loomis, J. M. (1982). Analysis of tactile and visual confusion matrices. *Perception & Psychophysics*, 31(1), 41-52. doi: 10.3758/BF03206199
- Loomis, J. M. (1993). Counterexample to the hypothesis of functional similarity between tactile and visual pattern perception. *Perception & Psychophysics*, 54(2), 179-184. doi: 10.3758/BF03211753
- Maidenbaum, S., Abboud, S., & Amedi, A. (2014). Sensory substitution: closing the gap between basic research and widespread practical visual rehabilitation. *Neuroscience & Biobehavioral Reviews*, 41, 3-15. doi: 10.1016/j.neubiorev.2013.11.007
- Mateeff, S., Popov, D., & Hohnsbein, J. (1993). Multi-aperture viewing: perception of figures through very small apertures. *Vision Research*, 33(17), 2563-2567. doi: 10.1016/0042-6989(93)90135-J
- Meijer, P. B. (1992). An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering*, 39(2), 112-121. doi: 10.1109/10.121642
- Michel, F., & Henaff, M. A. (2004). Seeing without the occipito-parietal cortex: Simultagnosia as a shrinkage of the attentional visual field. *Behavioural Neurology*, 15(1, 2), 3-13. doi: 10.1155/2004/836830
- Morgan, M. J., Findlay, J. M., & Watt, R. J. (1982). Aperture viewing: A review and a synthesis. *The Quarterly Journal of Experimental Psychology Section A*, 34(2), 211-233. doi: 10.1080/14640748208400837
- Nau, A., Bach, M., & Fisher, C. (2013). Clinical tests of ultra-low vision used to evaluate rudimentary visual perceptions enabled by the BrainPort vision device. *Translational Vision Science & Technology*, 2(3), 1-1. doi: 10.1167/tvst.2.3.1
- Norman, J. F., & Bartholomew, A. N. (2011). Blindness enhances tactile acuity and haptic 3-D shape discrimination. *Attention, Perception, & Psychophysics*, 73(7), 2323-2331. doi: 10.3758/s13414-011-0160-4

- O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5), 939-973. doi: 10.1017/S0140525X01000115
- Ortiz, T., Poch, J., Santos, J. M., Requena, C., Martínez, A. M., Ortiz-Terán, L., Turrero, A., Barcia, J., Nogales, R., Calvo, A., Martínez, J. M., Córdoba, J. L., & Pascual-Leone, A. (2011). Recruitment of occipital cortex during sensory substitution training linked to subjective experience of seeing in people with blindness. *PloS one*, 6(8), e23264. doi: 10.1371/journal.pone.0023264
- Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences*, 111(16), 6104-6108. doi: 10.1073/pnas.1322705111
- Pasqualotto, A., & Esenkaya, T. (2016). Sensory substitution: the spatial updating of auditory scenes “Mimics” the spatial updating of visual scenes. *Frontiers in Behavioral Neuroscience*, 10, 79. doi: 10.3389/fnbeh.2016.00079
- Pasqualotto, A., & Proulx, M. J. (2012). The role of visual experience for the neural basis of spatial cognition. *Neuroscience & Biobehavioral Reviews*, 36(4), 1179-1187. doi: 10.1016/j.neubiorev.2012.01.008
- Pasqualotto, A., Spiller, M. J., Jansari, A. S., & Proulx, M. J. (2013). Visual experience facilitates allocentric spatial representation. *Behavioural Brain Research*, 236, 175-179. doi: 10.1016/j.bbr.2012.08.042
- Pec, M., Bujacz, M., Strumillo, P., & Materka, A. (2008, September). Individual HRTF measurements for accurate obstacle sonification in an electronic travel aid for the blind. In *Signals and Electronic Systems, 2008. ICSES'08. International Conference on* (pp. 235-238). IEEE. doi: 10.1109/ICSES.2008.4673402
- Potisk, T. (2015). Head-Related Transfer Function. In *Seminar Ia, Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia*.
- Renier, L., Bruyer, R., & De Volder, A. G. (2006). Vertical-horizontal illusion present for sighted but not early blind humans using auditory substitution of vision. *Perception & psychophysics*, 68(4), 535-542. doi: 10.3758/BF03208756
- Renier, L., Laloyaux, C., Collignon, O., Tranduy, D., Vanlierde, A., Bruyer, R., & De Volder, A. G. (2005). The Ponzo illusion with auditory substitution of vision in sighted and early-blind subjects. *Perception*, 34(7), 857-867. doi: 10.1068/p5219
- Phillips, J. R., Johnson, K. O., & Browne, H. M. (1983). A comparison of visual and two modes of tactual letter resolution. *Perception & Psychophysics*, 34(3), 243-249. doi: 10.3758/BF03202952
- Röder, B., Teder-Sälejärvi, W., Sterr, A., Rösler, F., Hillyard, S. A., & Neville, H. J. (1999). Improved auditory spatial tuning in blind humans. *Nature*, 400, 162-166. doi: 10.1038/22106
- Perrott, D. R., & Saberi, K. (1990). Minimum audible angle thresholds for sources varying in both elevation and azimuth. *The Journal of the Acoustical Society of America*, 87(4), 1728-1731. doi: 10.1121/1.399421
- Rochlis, J. L. (1998). *A Vibrotactile Display for Aiding Extravehicular Activity (EVA) Navigation in Space* (Doctoral dissertation, Massachusetts Institute of Technology).
- Rodríguez-Hernández, A. F., Merino, C., Casanova, O., Modrono, C., Torres, M. Á, Montserrat, R., Navarrete, G., Burunat, E., & González-Mora, J. L. (2010). Sensory substitution for visually disabled people: computer solutions. *WSEAS Transactions on Biology and Biomedicine*, 7, 1-10.

- Ruggiero, G., Ruotolo, F., & Iachini, T. (2012). Egocentric/allocentric and coordinate/categorical haptic encoding in blind people. *Cognitive Processing*, 13(1), 313-317. doi: 10.1007/s10339-012-0504-6
- Ruggiero, G., Ruotolo, F., & Iachini, T. (2018). Congenital blindness limits allocentric to egocentric switching ability. *Experimental Brain Research*, 236(3), 813-820. doi: 10.1007/s00221-018-5176-8
- Sampaio, E., Maris, S., & Bach-y-Rita, P. (2001). Brain plasticity: 'visual' acuity of blind persons via the tongue. *Brain Research*, 908(2), 204-207. doi: 10.1016/S0006-8993(01)02667-1
- Spagnol, S., Baldan, S., & Unnthorsson, R. (2017, October). Auditory depth map representations with a sensory substitution scheme based on synthetic fluid sounds. In *Proceedings of the IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*. Luton, UK. doi: 10.1109/MMSP.2017.8122220
- Spanlang, B., Normand, J. M., Giannopoulos, E., & Slater, M. (2010, November). A first person avatar system with haptic feedback. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology* (pp. 47-50). ACM. doi: 10.1145/1889863.1889870
- Starkiewicz, W., & Kuliszewski, T. (1963). The 80-channel elektroftalm. In *Proceedings of the International Congress Technology Blindness, Am. Found. Blindness*, 1, 157.
- Stevens, J. C., & Choo, K. K. (1996). Spatial acuity of the body surface over the life span. *Somatosensory & Motor Research*, 13(2), 153-166. doi: 10.3109/08990229609051403
- Stoll, C., Palluel-Germain, R., Fristot, V., Pellerin, D., Alleysson, D., & Graff, C. (2015). Navigating from a Depth Image Converted into Sound. *Applied Bionics and Biomechanics*, 2015, 543492-543492. doi: 10.1155/2015/543492
- Striem-Amit, E., Guendelman, M., & Amedi, A. (2012). 'Visual' acuity of the congenitally blind using visual-to-auditory sensory substitution. *PloS one*, 7(3), e33136. doi: 10.1371/journal.pone.0033136
- Stronks, H. C., Mitchell, E. B., Nau, A. C., & Barnes, N. (2016). Visual task performance in the blind with the BrainPort V100 Vision Aid. *Expert Review of Medical Devices*, 13(10), 919-931. doi: 10.1080/17434440.2016.1237287
- Stronks, H. C., Walker, J., Parker, D. J., & Barnes, N. (2017). Training Improves Vibrotactile Spatial Acuity and Intensity Discrimination on the Lower Back Using Coin Motors. *Artificial Organs*. doi: 10.1111/aor.12882
- Strybel, T. Z., Manllgas, C. L., & Perrott, D. R. (1992). Minimum audible movement angle as a function of the azimuth and elevation of the source. *Human Factors*, 34(3), 267-275. doi: 10.1177/001872089203400302
- Urbanski, M., De Schotten, M. T., Rodrigo, S., Catani, M., Oppenheim, C., Touzé, E., Chokron, S., Méder, J-F., Lévy, R., Dubois, B., & Bartolomeo, P. (2008). Brain networks of spatial awareness: evidence from diffusion tensor imaging tractography. *Journal of Neurology, Neurosurgery & Psychiatry*, 79(5): 598-601. doi: 10.1136/jnnp.2007.126276
- Van Erp, J. B. (2005, March). Vibrotactile spatial acuity on the torso: effects of location and timing parameters. In *Eurohaptics Conference, 2005 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2005. World Haptics 2005. First Joint* (pp. 80-85). IEEE. doi: 10.1109/WHC.2005.144
- Vercillo, T., Tonelli, A., & Gori, M. (2018). Early visual deprivation prompts the use of body-centered frames of reference for auditory localization. *Cognition*, 170, 263-269. doi: 10.1016/j.cognition.2017.10.013

- Wacker, P., Wacharamanotham, C., Spelmezan, D., Thar, J., Sánchez, D. A., Bohne, R., & Borchers, J. (2016, May). VibroVision: An On-Body Tactile Image Guide for the Blind. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 3788-3791). ACM. doi: 10.1145/2851581.2890254
- Wan, C. Y., Wood, A. G., Reutens, D. C., & Wilson, S. J. (2010). Early but not late-blindness leads to enhanced auditory perception. *Neuropsychologia*, 48(1), 344-348. doi: 10.1016/j.neuropsychologia.2009.08.016
- Ward, J., & Meijer, P. (2010). Visual experiences in the blind induced by an auditory sensory substitution device. *Consciousness and Cognition*, 19(1), 492-500. doi: 10.1016/j.concog.2009.10.006
- Ward, J., & Wright, T. (2014). Sensory substitution as an artificially acquired synaesthesia. *Neuroscience & Biobehavioral Reviews*, 41, 26-35. doi: 10.1016/j.neubiorev.2012.07.007
- Weinstein, S. (1968). Intensive and extensive aspects of tactile sensitivity as a function of body part, sex, and laterality. In D. R. Kenshalo (Ed.), *The Skin Senses* (pp. 195-222). Springfield, IL: Thomas.
- Zwiers, M. P., Van Opstal, A. J., & Cruysberg, J. R. M. (2001). A Spatial Hearing Deficit in Early-Blind Humans. *Journal of Neuroscience*, 21(9), RC142 1-5.

Figure descriptions:

Figure 1. Spectrogram of noise at different elevations using head related transfer function, darker patches indicate higher amplitudes. Noise at lower elevations suppresses higher frequencies, while at higher elevations the higher frequencies are preserved while lower frequencies are suppressed.

Figure 2: Left image is of a Synaestheatre user, right image is of the Synaestheatre observing two disc stimuli that vary in distance to the sensor. In the visual condition, users directly observe the 16×8 depth map. The iPhone shows two identical 16×8 depth maps as each one is presented to separately to each eye within the VR headset, which is subjectively perceived as a single image. In the ‘auditory’ condition users do not receive any visual stimulation and only perceive sounds being played from those spatial locations are heard by the subject.

Figure 3. Left image shows a VibroVision vest user, the chest mounted camera provides 3D information that is convey via spatialised vibrations arranged in a 16×8 matrix (right image).

Figure 4. How stimuli are perceived by subjects. The top row showcases the depth condition and the bottom row showcases the verticality condition. The ‘Stimuli’ column shows visual image examples of stimuli disc locations. The ‘Visual SSD’ column showcases the visual images observed by subjects while wearing the VR headset, distance information is coded as pixel brightness (closer = brighter). The ‘Audio SSD’ column shows which audio files are active (audio icons) and their loudness (sound waves) to stimuli locations to the subject, distance information is coded as audio file loudness (closer = louder). Examples of the auditory SSD can be found at the following links: <https://youtu.be/QO2omDc3Orw> and https://youtu.be/5_rASaDsg5c. The ‘Haptic SSD’ column shows which vibrotactile motors would be active (white blocks) and their vibrotactile intensity (block size) to these stimuli locations, distance is coded as vibrotactile intensity. **As a result of visual angles, at certain thresholds (e.g. closer than 96.4 cm) closer stimuli would activate more ‘pixels,’ resulting in a larger area of stimulation across all SSDs.** Subjects were stationary in all conditions but could ‘look around’ by varying the pitch, yaw and roll of the sensor. Subjects were blindfolded in the auditory and haptic SSD conditions.

Figure 5: A topdown view of the depth condition. The (a) stimuli consists of two discs of 34 cm diameter suspended 100 cm from the floor, and 100 cm horizontally from one another. The task involved (b) systematically varying the stimuli in their distance (+/-) from the (c) central point indicated by the horizontal red dotted line, which is set to a distance of 120 cm from the (d) participants’ location. The participant is tasked with answering whether the left or right stimulus disc is closer to them using the provided sensory substitution device (Synaestheatre, VibroVision).

Figure 6: A first-person view of the verticality condition. The (a) stimuli consist of two discs of 34 cm diameter, positioned 120 cm in front of the participant’s location, the stimuli are set with a horizontal centre-to-centre distance of 100 cm from one another. The task involved (b) systematically varying the stimuli in their distance (+/-) from the (c) middle point indicated by the horizontal red dotted line. The participant is tasked with answering whether the left or right stimulus disc is higher using the provided sensory substitution device (Synaestheatre, VibroVision).

Figure 7: Mean confidence scores plotted against mean performance across all ten trials for vertical (top row) and depth (bottom row) perception for visual, haptic, and auditory SSDs respectively. Scores are given in blue, confidence in orange, error bars represent 95% confidence intervals.